R E P O R T R E S U M E S

ED 016 285 24 EA 001 054

A STUDY OF THE GENERALITY OF PRESENTING BEHAVIOR IN TEACHING.

BY- FORTUNE, JIMMIE C.
MEMPHIS STATE UNIV., TENN.

REPORT NUMBER BR-6-8464 PUB DATE 31 MAY 67
CONTRACT OEC-2-7-068464-0217
EDRS PRICE MF-$0.25 HC-$1.52 36P.

DESCRIPTORS- *TEACHER BEHAVIOR, *TEACHING METHODS, COMMUNICATION SKILLS, TEACHER EDUCATION, *ELEMENTARY SCHOOL TEACHERS, *ACADEMIC PERFORMANCE, COURSE CONTENT, TEACHER RATING, LITERATURE REVIEWS, SOCIAL STUDIES, MATHEMATICS, ENGLISH, PROGRAMED INSTRUCTION, *EFFECTIVE TEACHING, ELEMENTARY SCHOOL STUDENTS, STATISTICAL ANALYSIS, VIDEO TAPE RECORDINGS, MEMPHIS,

THIS STUDY IS CONCERNED WITH THE IDENTIFICATION AND ASSESSMENT OF SKILLS ASSOCIATED WITH TEACHER BEHAVIORS DIRECTED TOWARD THE PRESENTATION OF CONTENT. THE IDENTIFICATION OF THE VARIOUS SKILLS WAS MADE THROUGH OBSERVATION AND ISOLATION OF TEACHER VARIATIONS WHICH RESULT IN STUDENT PERFORMANCE DIFFERENCES ON PRECONSTRUCTED CRITERIA TESTS. THE GENERALITY OF TEACHER BEHAVIORS WAS ANALYZED FOR DIFFERENT GROUPS OF LEARNERS, DIFFERENT PACKETS OF CONTENT, AND A COMBINATION OF THESE TWO. AN ASSESSMENT STUDY WAS MADE TO ASCERTAIN THE RELATIONSHIP OF LEARNER PERFORMANCES TO SUPERVISORY RATINGS OF PARTICIPATING TEACHERS. ADAPTATIONS OF RYAN'S CLASSROOM OBSERVATION RECORD AND THE STANFORD TEACHER APPRAISAL GUIDE OF TEACHING COMPETENCE WERE USED TO GATHER THE SUPERVISORY RATINGS DATA. THE FINDINGS OF THE STUDY IMPLY GENERALITY OVER CONTENT PACKETS ACROSS LEARNER GROUPS, NO ASSUMPTIONS OF GENERALITY OVER LEARNER GROUPS ACROSS CONTENT PACKETS, AND NO GENERALITY OVER CONTENT PACKETS AND LEARNER GROUPS. THE PRINCIPAL IMPLICATION IS THAT ELEMENTARY TEACHERS SEEM TO PRESENT THE SAME PACKET OF CONTENT WITH SIMILAR SUCCESS TO DIFFERENT LEARNER GROUPS. NO SPECIFIC BEHAVIOR WAS FOUND TO GENERALLY DISCRIMINATE BETWEEN EXTREME TEACHERS, EVEN THOUGH SEVERAL COMMON DISCRIMINATING BEHAVIORS WERE FOUND BETWEEN ANY TWO OF THE THREE CONTENT AREAS. POSITIVE CORRELATIONS WERE FOUND BETWEEN SUPERVISORY RATINGS AND AVERAGE STUDENT PERFORMANCE SCORES. (HW)

ED016285

BR 6-8464
PA 24

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

BR FINAL REPORT
~~PROJECT~~ NO. 6-8464
CONTRACT NO. OEC2-7-068464-0217

A STUDY OF THE GENERALITY OF PRESENTING BEHAVIOR IN TEACHING

JUNE 1967

U. S. Department of
Health, Education, and Welfare

Office of Education
Bureau of Research

EA 001 054

A STUDY OF THE GENERALITY OF PRESENTING BEHAVIOR IN TEACHING

Project No. 6-8464
Contract No. OEC2-7-068464-0217

Dr. Jimmie C. Fortune

May 31, 1967

The research reported herein was performed pursuant to a Contract No. OEC2-7-068464-0217 with the Office of Education, U. S. Department of Health, Education, and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position or policy.

Memphis State University

Memphis, Tennessee

3. Contents

SECTION		PAGE
4.	Acknowledgments	iii
5.	Introduction	1
6.	Method	4
7.	Results	5
8.	Discussion	17
9.	Conclusions and Implications	28
10.	Summary	29
11.	References	31
	Eric Resume Form	33

TABLE		PAGE
1		6
2		7
3		8
4		9
5		10
6		12
7		13
8		14
9		15
10		16
11		18
12		19
13		20
14-A		21
14-B		22
15		23
16-A		24
16-B		25
16-C		26

4. Acknowledgments

My sincere appreciation goes to Mr. Howard Holst of WKNO Educational Television Cooperation for his advice and cooperation. Acknowledgment should also include the 1966-1967 class of the Mid-South Undergraduate Research Training Program and to Mr. Don England whose work was instrumental in the completion of this research.

5. Introduction

Background

The behavioral sciences have lagged in the production of methods of adequate, objective evaluation and assessment of classroom teaching. This slowness of development is due to the complexity of the variables with which the behavioral scientists must work. Because many of these variables are of a subjective nature there exists problems in measurement and instrumentation. The multivariate characteristics of behavioral science variables complicate the identification, isolation, and assessment of linear effects of any one variable. The changing nature of classroom practice and its relationship to cultural evolution and values promote the necessity of subjective and psychophysical judgments and reduce the chances of reproducibility. The solution to classroom analysis may become reality if a single teaching encounter can be subjected to several different analyses.

The use of video-tape to make teaching encounters an item of record is a valuable research tool. The recording of teaching encounters enables trained observers to analyze the relationships of teacher behaviors to student performances and other variables in the classroom from several points of view. This strategy involves focusing upon a specific teacher behavior directed toward the accomplishment of a specific classroom goal, video-tape recording several incidents of the teacher behavior, assessing student levels of learning, and then isolating behavioral incidents resulting in minimum and maximum goal attainment for cross comparison. From repeated applications of this strategy, generality of teacher behaviors and evidence of successful practice can be subjected to analysis after statistical corrections of the criteria measures have been for student differences.

Problem

In the inventory of tools routinely used by the classroom teacher there is an intricate array of communication skills and techniques whose focal point is the presenting of curricular content. An understanding of these behaviors is of great importance in adequately preparing classroom teachers and in the improvement of classroom practices. What special skills should a teacher-trainee develop in order to insure his success in presenting curricular content? Can these skills be isolated so that they may be explored and described for teacher-trainees in teacher preparation programs?

The problem of concern in this study was the identification and assessment of particular skills associated with teacher behavior in presenting content to students in grades four, five, and six. Identification of these skills was made through the observation and isolation of teacher variations which resulted in student performance differences on preconceived criteria tests. A study of the generality of the teacher behaviors over different groups of students, different packets of content, and a combination of the two was made to provide a basis for investigation of variations in teacher behavior. A study of the assessment problem was made through the investigation of relationships between supervisory rating of teacher performances, made on adaptations of two commonly used rating scales, and adjusted means of student test performances using student performances on a programmed learning packet as a covariant.

Related Literature

Presenting behavior is defined here to include teacher actions directed toward the induction of learner set, the transmission of instructional goals, the orientation of the learner to skills required for achievement of instructional goals, the exposition of instructional information, and the rendation of task direction. Teacher presentation has been operationally defined as a method of instruction (16) and was included in an analysis of the teaching act as far back as Herbart (5). The Stanford Teacher Eduation Program has for the past three years included experiences in aspects of presenting behavior in intern-teacher preparation. Empirical evidence generated in the study of micro-teaching, a teacher-training construct developed at Stanford, indicates that both set induction (3,7,2, and 6) and lecturing behavior (1) are teaching skills which produce measurable differences in learner perceptions of teaching performances. A survey of studies designed to contrast lecture and discussion methods of teaching indicates that lecturing is a defensible method of teaching as shown in several studies of the two methods (11).

The proposed design for this study is an attempt to take advantage of past attempts to ascertain teacher competence. In a similar design, a previous study of explaining behavior submitted for presentation at the AERA Chicago conference revealed some generality of explaining behavior over packets of content and small but significant (PL .01) positive correlations to items on the Stanford Appraisal Guide of Teacher Competence (9). In the past, four basic techniques have been used in attempts to solve the criterion problem: (1) to identify characteristics of the "effective teachers;" (2) to measure pupil behavior change; (3) to establish norms for effective teachers on a series of standardized measures of personality variables, attitudes, and academic skills;

and (4) to rate teachers on actions logistically connected with competent teaching. Perhaps, the best example of the first attempt is the study of Ryans (15). McCall (12) in his work with the merit pay system was perhaps the most successful investigator to overcome the influencing effects of the environment and to isolate the effects of a specific teacher. The Wisconsin studies by Barr (4) and the work of Medley and Mitzel (13) illustrate technique three. The Purdue Scale studied by Remmers (14) and the Stanford Appraisal Guide of Teaching Competence (10) are examples of the fourth technique.

An earlier study done by Central Midwestern Regional Educational Laboratory on "The Generality of Presenting Behaviors in Teaching Preschool Children (8) implied the existence of generality of presenting behavior over content packets across learner groups. There was some generality of presenting behavior over learner groups across content packets, however this generality produced small corrections and was considered unstable. Four behaviors were identified as discriminating factors of preschool teacher success. They were (1) the teacher made provisions for the children to explore and manipulate important elements of the content, (2) the teacher introduced and provided for opportunities for student verbal practice, (3) the teacher reinforced correct responses, and (4) the teacher appeared sensitive to the children's needs, interests, and questions.

Objectives

This study has three principal purposes. First, the study was for the investigation of the generality of presentation effectiveness as seen over groups of learners in the fourth, fifth, and sixth grades, over packets of content consisting of social studies, mathematics, and English materials, and over a combination of learner groups and content packets. Second, the study was designed to analyze and evaluate teacher actions in order to isolate variations in teacher behavior which produce differences in learner performances on examinations constructed to measure comprehension of the content packets. Third, supervisory ratings of the participating teachers were used to determine the relationship of learner performances to these ratings. These supervisory ratings were made by two independent raters, one of whom was the supervising teacher whose class was involved, the other rater was a student currently enrolled in the School of Education, Memphis State University. Each supervisor rated the teacher on an adaptation of both the adjective checklist devised by Ryans and on the Stanford Appraisal Guide of Teacher Competence. To reduce the effects of pupil differences this study used comparable programmed content packets as a baseline for the adjustment of individual performance.

6. Method

Procedures

Each teacher's class in grades four, five, and six was randomly divided into two learner groups (group 1 and group 2). The teacher was asked to teach three lessons of a four lesson unit (A, B, C, and D) to either or both of the learner groups. During the first instructional encounter, the teacher taught lesson A to group 1, while group 2 received lesson B in a programmed learning packet. The second encounter had the teacher teaching lesson B to group 1 while group 2 received lesson D in programmed form. In the third encounter group 2 was taught lesson A and group 1 was given programmed lesson C. The final encounter had the teacher teaching lesson C to group 2 while group 1 received lesson D in programmed form. The entire sequence may be seen in the following diagram:

	Group 1	Group 2
Encounter 1	Taught A	Programmed B
Encounter 2	Taught B	Programmed D
Encounter 3	Programmed C	Taught A
Encounter 4	Programmed D	Taught C

Several preparations were necessary in order to gather the data necessary for the accomplishment of the research goals. These preparations included the construction of the lessons to be presented, the building of student performance tests on each lesson, the selection and revision of the supervisory rating scales, and the determination of which lessons to be taught and which to be presented in programmed-instruction format.

The programmed lessons were constructed parallel to the taught lessons and provided the necessary data for the establishment of covariants to handle statistically learner group differences and content packet difficulty. The teacher presentations of each of the four encounters was recorded on video-tape and saved for the analysis of group deviations and teacher variations.

Sample

Forty-two teaching encounters including 15 in English, 14 in mathematics, and 13 in social studies were taught by teacher trainees in the Memphis State University School of Education. The teacher trainees who participated in the study were juniors or seniors currently enrolled in the School of Education though not necessarily pursuing courses of study leading to teacher certification. None had previous professional experience.

The learner groups who participated in the study were fourth, fifth, and sixth grade students at the Campus School, Memphis State University. The average class size was approximately twenty-eight providing learner groups of approximately fourteen each, since a random dividing of the classes was included in the general design. In total 256 pupils participated in the experiment, with each class receiving instruction in social studies, mathematics, and English.

Eight units of content were developed in each of three areas, namely, mathematics, social studies, and English. The mathematics units included: fractions, ratio, per cent, decimals, bases other than ten, elementary number theory, and lattice squares. In social studies the concepts advanced were: a contrast of U.S.S.R. and United States governments, the three branches of U. S. government, an historical-geographical study of the British Isles, election in the United States, the party system of politics, and the democratic system. The English topics taught were: general reference books, types of sentences, sentence structure, comparing adjectives and adverbs, sentence patterns and transformational grammar, and use of the library.

Inherent in the research design is the utilization of programmed learning packets as a baseline for the comparison of the teacher presentations in respect to groups differences in ability to respond. Each concept presentation, whether programmed or taught, was limited to no less than ten nor more than fifteen minutes. The teacher presentations were made in the audio-visual room of the Campus School, Memphis State University. The programmed packets were presented to the students in the classroom usually used by the students in the Campus School.

The evaluation of the fourth, fifth, and sixth grade students was accomplished by means of a short answer paper and pencil test over a combination of the four concepts presented during the preceding hour. The various tests were uniformly based upon a format of twelve short answer questions.

7. Results

The Generality Study

The central research objective of the study was the ascertainment of generality of teacher effectiveness in presentation behaviors directed toward elementary school children. As part of the research design an attempt was made to identify differences in generality that may be explained through content differences as would be seen in the teaching of English, mathematics, and social studies.

Pupil performance was measured by twelve-item, short answer tests developed to evaluate each content packet. The tests for the content packets were developed in the Mid-South Undergraduate Research Training Program and were subjected to a split-half reliability study. Table 1 indicates the reliability correlations found for the six-sets of content tests. Although these coefficients were disappointingly low, probably due to the small number of items on each test, additional reliability was gained through the utilization of the mean of the learner groups which were composed of twelve to sixteen pupils.

Table 1

Split-half Reliability of Six Sets of Content Tests
N=30 on each set

Content	Split-half Correlation	Significance Level
Mathematics Test 1	.64	.01
Mathematics Test 2	.61	.01
Social Studies Test 1	.58	.01
Social Studies Test 2	.70	.01
English Test 1	.78	.01
English Test 2	.73	.01

The scores of the learner groups on the teacher taught lessons were adjusted by the corresponding scores made on the parallel programmed instruction lessons. The adjustment of the scores was used to minimize learner group differences and content packet differences. A generality study was made for each of the three content areas and an overall analysis was made on the entire sample of teachers in the study.

Product-moment correlation coefficients were used to estimate generality of presenting behavior over groups of students, over packets of content and over groups of students and content packets. Tables 2, 3, and 4 report the matrices obtained for the content areas and Table 5 shows the correlation matrix for the overall generality study.

Table 2

Correlation Matrix for English

Matrix of Product-Moment Correlations Computed on Adjusted
Performance Scores of Fifteen Teachers Presenting English Content

Packets to Two Learner Groups Each

Teaching Encounter	Packet A to Group 1 A-1	Packet B to Group 1 B-1	Packet A to Group 2 A-2	Packet C to Group 2 C-2
A-1	1.00	-.27	.57	-.07
B-1		1.00	-.35	.53
A-2			1.00	.12
C-2				1.00

Table 3

Correlation Matrix for Mathematics

Matrix of Product-Moment Correlations Computed on Adjusted
Performance Scores of Fourteen Teachers Presenting Mathematics Content

Packets to Two Learner Groups Each

Teaching Encounter	Packet A to Group 1 A-1	Packet B to Group 1 B-1	Packet A to Group 2 A-2	Packet C to Group 2 C-2
A-1	1.00	-.17	.68	.05
B-1		1.00	-.29	.02
A-2			1.00	.04
C-2				1.00

Table 4

Correlation Matrix for Social Studies

Matrix of Product-Moment Correlations Computed on Adjusted
Performance Scores of Thirteen Teachers Presenting Social Studies Content

Packets to Two Learner Groups Each

Teaching Encounter	Packet A to Group 1 A-1	Packet B to Group 1 B-1	Packet A to Group 2 A-2	Packet C to Group 2 C-2
A-1	1.00	-.14	.70	-.45
B-1		1.00	-.42	.42
A-2			1.00	-.43
C-2				1.00

Table 5

Correlation Matrix for Total Sample

Matrix of Product Moment Correlations Computed on Adjusted
Performance Scores of Forty-two Teachers Presenting Content

Packets to Two Learner Groups Each

Teaching Encounter	Packet A to Group 1 A-1	Packet B to Group 1 B-1	Packet A to Group 2 A-2	Packet C to Group 2 C-2
A-1	1.00	-.25	.69	-.11
B-1		1.00	-.31	.65
A-2			1.00	.03
C-2				1.00

Tape Analysis

As part of the research design an attempt was made to gain repeatable analyses of live teaching through the utilization of video-tape recordings. The analysis of the tapes did not include any existing coding systems or appraisal schedules. Three judges were employed to develop a format of analysis from behaviors they observed on three random tapes. The three judges were Dr. William Johnson and Dr. Norman Dodl of the University of Illinois and Dr. Don Johnston of Memphis State University, each of whom is involved in teacher training and television utilization at their respective institutions. From repeated observations of the three random tapes a form for the analysis of extreme tapes was developed.

A composite guide of sixteen basic characteristics was selected for the analysis of the tapes. Twenty-eight characteristics were named by only one judge or were named for only one or two teachers. The sixteen characteristics were then given the judges and a reliable method of coding teacher behaviors was devised.

Table 6 indicates the sixteen basic characteristics identified to form the category system for tape analysis.

For each content unit the three highest, three lowest and two mean teachers in regard to their mean increment of teaching success were selected. The three judges reviewed the 48 teaching episodes in random order and on the basis of categories developed previously. The three judges further described the teaching in terms of characteristic behavior and expected pupil performance.

Tables 7, 8, and 9 indicate the amount of agreement of the three independent judges on the rating of the extreme teachers in English, social studies and mathematics respectively. As can be seen on these tables agreement was generally high between the judges. In the analysis of the six extreme English teachers over the sixteen characteristics the judges were in total agreement on 65 cases, two judges agreed in 30 cases and the element of disagreement was one between quality of characteristic in 28 of the 30 cases, and the judges disagreed in only one incident. In the analysis of the six extreme social studies teachers over the sixteen characteristics the judges were in total agreement on 66 cases, two judges agreed on 27 cases with 21 of these 27 being disagreements in quality of the characteristic, and the judges disagreed on 3 cases. In the analysis of the six extreme mathematics teachers over the sixteen characteristics the judges were in total agreement on 56 cases, two judges agreed on 37 cases which included 26 quality disagreements by one of the judges, and the judges were in disagreement on 3 cases.

Table 6 Tape Review Format

Tape #_____ Teacher #_____

		Ratings			Comments
		++	+	-	
1. Introduction of set	a. instructional				
	b. rapport				
2. Lesson Development	a. principle points				
	b. discriminations				
	c. correct label				
	d. review/repetition				
	e. association (experimental ref.)				
3. Interaction	a. actual (verbal)				
	b. implicit (non-verbal)				
4. Reinforcement					
5. Discussion Skills	a. probing				
	b. patience to wait for response				
	c. integrates response into lesson				
6. Acceptance of pupil					
7. Self assurance, poise					
8. Lesson completion/summary					

Table 7*

Agreement Frequencies of Judges' Ratings
of English Teachers on the Sixteen Behaviors

		High 3 Teachers			Low 3 Teachers		
		All Agree	Two Agree	All Disagree	All Agree	Two Agree	All Disagree
1.	a.	3			1	2	
	b.	3				3	
2.	a.	3			2		1
	b.	2	1		1	2	
	c.	3			1	2	
	d.	3			2	1	
	e.	2	1		3		
3.	a.	2	1		1	2	
	b.	2	1		3		
4.		3			3		
5.	a.	1	2		3		
	b.	2	1		2	1	
	c.	1	2		1	2	
6.		2	1		2	1	
7.		3			2	1	
8.		1	2		2	1	

*Numbered columns refer to categories shown on Table 6, Tape Review Format, p. 12.

Table 8*

Agreement Frequencies of Judges' Ratings
of Social Studies Teachers on the Sixteen Behavi

		High 3 Teachers			Low 3 Teachers		
		All Agree	Two Agree	All Disagree	All Agree	Two Agree	All Disagree
1.	a.	2	1		2	1	
	b.		3		1	2	
2.	a.	2	1		1	2	
	b.	3			1	1	1
	c.	3			2	1	
	d.	3			3		
	e.	1	2		3		
3.	a.	2	1		3		
	b.	3			2	1	
4.		3			2	1	
5.	a.	1	2		3		
	b.	3			2	1	
	c.	3			2	1	
6.		3				2	1
7.		3				2	1
8.		2	1		2	1	

*Numbered columns refer to categories shown on Table 6, Tape Review Format, p. 12.

Table 9*

Agreement Frequencies of Judges' Ratings
of Mathematics Teachers on the Sixteen Behaviors

		High 3 Teachers			Low 3 Teachers		
		All Agree	Two Agree	All Disagree	All Agree	Two Agree	All Disagree
1.	a.		3		2	1	
	b.	3			2	1	
2.	a.	1	2		1	2	
	b.	1	2		1	2	
	c.	1	2		1	1	1
	d.	2	1			3	
	e.	3			2		1
3.	a.	3			2	1	
	b.	3			2	1	
4.		2	1			2	1
5.	a.	2	1		3		
	b.	3			2	1	
	c.	2	1		1	2	
6.		3			2	1	
7.		3				3	
8.		2	1		1	2	

*Numbered columns refer to categories shown on Table 6, Tape Review Format, p. 12.

Table 10

The Overall Ratings of the Extreme Teachers by the Three Judges and the Average Adjusted Performance Scores

Teacher	Judge 1	Judge 2	Judge 3	Average Performance Scores
English Teacher 1	upper	upper	upper	upper
English Teacher 2	middle	middle	lower	middle
English Teacher 3	upper	upper	upper	upper
English Teacher 4	lower	lower	lower	lower
English Teacher 5	upper	upper	upper	upper
English Teacher 6	lower	lower	lower	lower
English Teacher 7	upper	middle	middle	middle
English Teacher 8	lower	middle	middle	lower
Mathematics Teacher 1	upper	upper	upper	upper
Mathematics Teacher 2	upper	upper	middle	upper
Mathematics Teacher 3	middle	lower	middle	middle
Mathematics Teacher 4	lower	lower	lower	lower
Mathematics Teacher 5	lower	middle	lower	lower
Mathematics Teacher 6	upper	middle	upper	upper
Mathematics Teacher 7	lower	lower	lower	lower
Mathematics Teacher 8	middle	upper	upper	middle
Social Studies Teacher 1	upper	middle	middle	middle
Social Studies Teacher 2	upper	upper	upper	upper
Social Studies Teacher 3	upper	middle	upper	upper
Social Studies Teacher 4	middle	lower	lower	lower
Social Studies Teacher 5	lower	upper	middle	middle
Social Studies Teacher 6	middle	lower	lower	lower
Social Studies Teacher 7	lower	upper	upper	upper
Social Studies Teacher 8	lower	lower	lower	lower

Since one of the research goals of the study was to identify effective teacher behaviors in presenting content, positive incidents of each characteristic of the tape analysis format were tallied by the judges for each teacher. Tables 11, 12, and 13 indicate the frequency of response tallies for each group of extreme teachers. As can be observed from these tables some categories for each content area seemed to discriminate between the extreme teachers.

Assessment Study

For the assessment study the two scales selected were (1) an adjective checklist similar to the classroom observation record of Ryans and (2) the Stanford Teacher Appraisal Guide of Teaching Competence. Both scales had to be adapted to the brief teaching encounters and the planning limitations placed on the teacher by the curriculum prescriptions made by the programmed learning packets.

Each teacher was rated on each scale by two independent supervisors after each teaching encounter. Product-moment correlations for each content area were computed between the adjusted performance scores of the teachers and the items on the appraisal scale. Table 14 indicates the matrix obtained for the adjective checklist and Table 15 indicates the correlation matrix for the Stanford Appraisal Guide.

An interitem correlation matrix was computed between the two scales to ascertain the reliability of the two sets of ratings. Table 16 reports this matrix computed on the 42 teachers.

8. Discussion

Generality Study

There was a marked similarity of the four matrices obtained in the generality study. The coefficients designated to estimate generality over content packets across learner groups are shown on the tables as the correlations between A-1 to A-2. These correlations were significantly (P .01) positive and indicated stable generality over each of the content areas and over the total analysis. The correlations designated to estimate generality over learner groups across content packets are shown on the tables as correlations between A-1 and B-1 and between A-2 and C-2. The coefficients show some consistency results and indicate perhaps a time-factor could have intervened causing the three coefficients to estimate different kinds of generality over content packets and over learner groups. The coefficients computed between A-1 and C-2 and Between B-1 and A-2 were negative and small indicating no generality. The coefficients computed between B-1 and C-2 produced positive correlations

Table 11*

Frequency of Behaviors Tallied
by the Three Judges on the Extremes of the English Teachers

		3 Highest			3 Lowest		
		++	+	-	++	+	-
1.	a.	3-3	3		1	1-3-2	2
	b.	3	3-3			1-1-1	2-2-2
2.	a.	3-3-3			1	1-3	3-1
	b.	1-3	2-3		1	2-2	3-1
	c.	3-3-3			2-2	1-1	3
	d.	3-3-3				1	2-3-3
	e.	3	2-3	1			3-3-3
3.	a.	2-3	3-1			1-1	2-2-3
	b.	2-3	3-1				3-3-3
4.		4	3-3				3-3-3
5.	a.	1-3	2-2				3-3-3
	b.	1-3	3-2			2	3-1-3
	c.	1-1-3	2-2			1-1	2-2-3
6.		1-3-3	2			3-1	3-2
7.		3-3-3			2	1	3-3
8.		2-1-3	1-2		1	2	3-3

*Numbered columns refer to categories shown on Table 6, Tape Review Format, p. 12.

Table 12*

Frequency of Behaviors Tallied by the Three Judges
on the Extremes of the Social Studies Teachers

		3 Highest			3 Lowest		
		++	+	-	++	+	-
1.	a.	1	3-2-3			2-3-3	1
	b.	2-1	1-1-2	2	1	1	3-2-2
2.	a.	1-3-3	2		1-2	2-1-3	
	b.	3-3	3		2-1	1-1	3-1
	c.	3	3-3		3	2-3	1
	d.	3-3	3		3	3	3
	e.	1-2	2-1-3			3	3-3
3.	a.	3	3-1	2		3-3	3
	b.	3	3	3		2	3-3-1
4.			3-3-3			2-3	1-3
5.	a.	1	2-1	2-3			3-3-3
	b.	3	3	3		3-2	3-1
	c.		3	3-3		1	2-3-3
6.		3	3-3		1	2-1-1	1-2-1
7.		3	3-3		2-1-2	1-1-1	1
8.		3	3-2	1	2	3	1-3

*Numbered columns refer to categories shown on Table 6, Tape Review Format, p. 12.

Table 13 *

Frequency of Behaviors Tallied by the Three Judges
on the Extremes of the Mathematics Teachers

		3 Highest			3 Lowest		
		++	+	-	++	+	-
1.	a.	1-1	2-2-2	1		2	3-3-1
	b.	3	3	3	1	2	3-3
2.	a.	2-2-1	1-1-2			2-2	3-1-1
	b.	3-2-1	1-2		1	3-2-2	1
	c.	3-1	2-2	1	1	3-1-1	2-1
	d.	3-3-2	1			2-1-2	1-2-1
	e.	3	3	3	1	1	3-3-1
3.	a.	3	3-3		1	3-3-2	
	b.	3	3	3	1	2	3-3
4.		3	3-1	2	1	1-2-1	2-1-1
5.	a.	1	2-3	3			3-3-3
	b.		3-3-3			2-3-3	1
	c.	1	2-3-3			1-1	3-2-2
6.		3-3		3	2	3-1	3
7.		3-3	3		1	2-2-2	1-1
8.		3	3-1	2		1-2	3-2-1

*Numbered columns refer to categories shown on Table 6, Tape Review Format, p. 12.

Table 14-A

Correlations Between Average Performance Scores
and Supervisory Ratings Made on the Adjective Check-list

	Social Studies Teachers N=13	Mathematics Teachers N=14	English Teachers N= 15
Partial Fair (1)	.68	.64	.57
Autocratic Democratic (2)	.66	.64	.54
Aloof Responsive (3)	.66	.59	.48
Restricted Understanding (4)	.64	.58	.40
Harsh Kindly (5)	.68	.65	.51
Dull Stimulating (6)	.63	.62	.52
Stereotaped Original (7)	.67	.56	.50
Apathetic Alert (8)	.62	.52	.55
Unimpressive Attractive (9)	.64	.61	.57

Table 14-B

Correlations Between Average Performance Scores
and Supervisory Ratings Made on the Adjective Check-list

	Social Studies Teachers N=13	Mathematics Teachers N=14	English Teachers N=15
Evading Responsible (10)	.65	.61	.49
Erratic Steady (11)	.68	.60	.55
Excitable Poised (12)	.62	.65	.56
Uncertain Confident (13)	.61	.64	.50
Disorganized Systematic (14)	.54	.63	.49
Inflexible Adaptable (15)	.57	.54	.43
Pessimistic Optimistic (16)	.30	.51	.34
Immature Integrated (17)	.52	.53	.42
Narrow Broad (18)	.46	.32	.41

Table 15

Correlations Between Average Performance Scores
and Supervisory Ratings Made On
The Stanford Teacher Appraisal Guide of Teaching Competence

	Social Studies Teachers N=13	Mathematics Teachers N=14	English Teachers N=15
Statement of Goals (1)	.59	.57	.49
Organization of Lesson (2)	.67	.69	.66
Beginning the Lesson (3)	.71	.69	.60
Lesson Presentation (4)	.73	.71	.71
Attending Behavior (5)	.69	.50	.64
Pacing the Lesson (6)	.62	.63	.59
Rapport (7)	.58	.63	.59
Ending the Lesson (8)	.69	.63	.68

Table 16-A

Correlations Made Between Supervisory Ratings Made on the Adjective Checklist and Supervisory Ratings Made on the Stanford Teacher Appraisal Guide of Teaching Competence (N=42)

ITEM	1 Partial Fair	2 Autocratic Democratic	3 Aloof Responsive	4 Restricted Understanding	5 Harsh Kindly	6 Dull Stimulating
1 Statement of Goals	.31	.37	.35	.21	.33	.30
2 Organization of Lesson	.58	.61	.67	.70	.57	.50
3 Beginning the Lesson	.36	.40	.41	.43	.38	.48
4 Lesson Presentation	.60	.58	.58	.63	.73	.53
5 Attending Behavior	.21	.32	.26	.32	.51	.47
6 Pacing the Lesson	.51	.49	.61	.58	.54	.61
7 Rapport	.41	.37	.37	.36	.61	.48
8 Ending the Lesson	.30	.22	.28	.39	.63	.56

Table 16-B

Correlations Made Between Supervisory Ratings Made on the Adjective Checklist and Supervisory Ratings Made on the Stanford Teacher Appraisal Guide of Teaching Competence (N=42)

ITEM	7 Stereotyped Original	8 Apathetic Alert	9 Unimpressive Attractive	10 Evading Responsible	11 Erratic Steady	12 Excitable Poised
1 Statement of Goals	.28	.43	.28	.29	.31	.23
2 Organization of Lesson	.49	.53	.65	.65	.68	.51
3 Beginning the Lesson	.47	.31	.40	.47	.53	.30
4 Lesson Presentation	.69	.67	.77	.65	.67	.63
5 Attending Behavior	.51	.28	.73	.38	.47	.40
6 Pacing the Lesson	.48	.41	.58	.34	.36	.38
7 Rapport	.36	.53	.48	.38	.38	.31
8 Ending the Lesson	.61	.31	.64	.57	.59	.41

Table 16-C

Correlations Made Between Supervisory Ratings Made on the Adjective Checklist and Supervisory Ratings Made on the Stanford Teacher Appraisal Guide of Teaching Competence (N=42)

ITEM	13 Uncertain Confident	14 Disorganized Systematic	15 Inflexible Adaptable	16 Pessimistic Optimistic	17 Immature Integrated	18 Narrow Broad
1 Statement of Goals	.46	.43	.28	.12	.41	.03
2 Organization of Lesson	.51	.70	.64	.18	.29	.43
3 Beginning the Lesson	.59	.50	.43	.36	.39	.29
4 Lesson Presentation	.56	.67	.55	.41	.38	.41
5 Attending Behavior	.50	.28	.15	.09	.51	.16
6 Pacing the Lesson	.46	.41	.53	.48	.37	.32
7 Rapport	.57	.51	.09	.51	.44	.09
8 Ending the Lesson	.34	.39	.18	.27	.42	.22

adequate to imply generality. In the light of the conflicting results in the difference of the correlations which were consistent over the three content areas and also in the total sample, no generality can be assumed to exist over content packets and over learner groups.

Tape Analysis

There was some concern on the part of the three judges that the tape analysis format was inadequate to cover the dimensions of behavior shown on every tape. This concern was of merit in specific instances; however, many such special behaviors are difficult to categorize and often appear to be unique to individuals. The tape review format although inadequate to totally describe the presentation methodology present on all of the tapes does however capture most of the salient features of presenting behavior.

From the tape analysis of extreme groups the judges were able to identify several discriminating factors which were usually present in the outstanding teachers and usually absent in the less effective extremes. These behaviors will be stated for each subject area.

The judges ratings shown on Tables 11, 12, and 13 indicate the following discriminating factors between the extreme teachers:

Of the English teachers; (1) Actual (verbal) interaction, (2) implicit (non-verbal) interaction, (3) reinforcement, (4) probing skills, (5) patience to wait for response, (6) integration of student response into lesson, and (7) acceptance of pupil.

Of the Social Studies teachers; (1) Instructional set, (2) review and repetition, (3) reinforcement, and (4) acceptance of pupil.

Of the Mathematics teachers; (1) Instructional set, (2) lesson development, principle points, (3) review and repetition, (4) patience to wait for response, and (5) integration of student response into lesson.

Unfortunately, no behaviors appeared generalizable over all three content areas. This may explain why univereal efforts toward teacher instruction often result in little empirical support of teacher education programs.

Assessment Study

Table 16 of the assessment study produced a matrix generally composed of small but positive correlations between the two sets of supervisory ratings on the two different scales. A matrix of this nature seems to indicate some validity to the trained observer of teaching in that the two observers using different structured guides were in some agreement upon how they evaluated the teaching performances.

The correlations between the adjusted performance scores and the adjective checklist were generally positive and sufficiently large despite the small group samples. There was unfortunately much similarity of each adjective in its relationship to the average performance scores. This similarity seems to indicate some failure on the part of the supervisor to differentiate item-measuring and to make global ratings toward some basic value which may be identified as "good teaching."

The correlations between adjusted performance scores and the Stanford Teacher Appraisal Guide items were also generally positive and similar to the correlations found between adjusted performance scores and the adjective checklist. Again the items produced similar correlations indicating a possibly global interpretation of item meaning on the part of the supervisors. Item 5 on lesson presentation did produce generally higher correlations indicating some discriminate validity of the scale. Surprisingly, the content areas did not seem to operate as consistent variables in the assessment of teaching performances.

9. Conclusions and Implications

The coefficients designated to estimate generality over content packets across learner groups indicated stable generality over each of the three content areas and over the total analysis. Elementary teachers seem to present the same packet of content with similar success to different learner groups. The correlations designated to estimate generality over learner groups across content packets show some consistency over the content areas, however, small and negative coefficients preclude any assumption of generality. The correlations designated to estimate the generality of presenting behavior over content packets and over learner groups reveal inconsistent results upon analysis and imply a time-factor variable could have intervened causing the three coefficients to estimate different kinds of generality over content packets and over learner groups. Since there exist conflicting results in the different correlations over the three content areas and in the total sample, no generality can be implied over content packets and learner groups.

The only significant generality that can be implied from this study is generality over content packets.

The tape analysis provides some explanation of this lack of generality in that there was a difference of discriminating behaviors over content areas. This difference of discriminating behaviors could well exist within content areas as well as across content areas. Although there was a general overlap of discriminating behaviors over any two content groups, no specific behavior proved to discriminate between successful and unsuccessful teachers over all three content areas.

The consistency of the judges in the tape analysis and the positive correlations obtained between the supervisory ratings and the average performance scores indicate that within structured conditions educators can identify and agree upon components of "good teaching." The simplicity of the behavior and the more concise definition of presentation may be an explanation of the increase in correlations obtained in this study as compared to those obtained in the earlier study of explaining behaviors.

10. Summary

This study is concerned with the identification and assessment of skills associated with teacher behaviors directed toward the presenting of content. This is necessary to the continuing improvement of teacher education programs in their developing methods by which educators can make objective evaluations and assessments of classroom teaching practices.

The identification of the various skills was made through observation and isolation of teacher variations which result in student performance differences on preconstructed criteria tests. The study analyzed the generality of teacher behaviors over different groups of learners, different packets of content, and a combination of the two.

An assessment study was made to ascertain the relationship of the learner performances to supervisory ratings of the participating teachers. Adaptations of Ryan's Classroom Observation Record and the Stanford Teacher Appraisal Guide of Teaching Competence were used to gather the supervisory ratings data.

The findings of the study imply generality over content packets across learner groups, no assumptions of generality over learner groups across content packets, and no generality over content packets and learner groups. The principal implication suggests that elementary teachers seem to present the same packet of content with similar success to different learner groups. No specific behavior

was found to generally discriminate between extreme teachers, even though several common discriminating behaviors were found between any two of the three content areas.

Positive correlations were found between supervisory ratings and average student performance scores. Although several of these correlations could not be assumed different from zero, the agreement of the supervisors on the two scales was significant.

11. References

1. Allen, Dwight W., Cooper, James, and Fortune, Jim C. An Analysis of the 1965 Summer Microteaching Clinic. Unpublished report, Stanford Teacher Education Program, Stanford University, Stanford, California. August 1965. 4p.

2. Allen, Dwight W. and Fortune, Jim C. An Analysis of Microteaching; A New Procedure in Teacher Education. A paper read to the 1965 American Education Research Association Annual Convention; Chicago, Illinois. February 1965. 3p.

3. Aubertine, Horace. An Experiment in the Set Induction Process and Its Application in Teaching. Unpublished dissertation, Stanford University, Stanford, California. August 1963. p. 22.

4. Barr, A. S. "The Measurement and Prediction of Teaching Efficiency: A Summary of Investigation," Journal of Experimental Education. 16, 1948. p. 203-283.

5. Broudy, Harry S. "Historic Exemplars of Teaching Method," Handbook of Research on Teaching edited by N. L. Gaye. Chicago, Illinois: Rand McNally & Company, 1965. p. 1-43.

6. Bush, Robert N. and Allen, Dwight W. Micro-Teaching. A report on an address given by Bush to the Santa Barbara Conferences on Teacher Education, 1964: Stanford University, Stanford, California. p. 2.

7. Fortune, Jim C. Instructional Set, Cognitive Closure, and Test Anxiety in the Presentation of Social Studies Content. Unpublished dissertation; Stanford University, Stanford, California. August 1965. p. 56-63.

8. Fortune, Jim C. The Generality of Presenting Behaviors in Teaching Preschool Children. Unpublished report, Central Midwestern Regional Education Laboratory, St. Ann, Missouri. August 1966. p. 75-76.

9. Fortune, Jim C., Gage, N. L., and Shutes, Robert. The Generality of Explaining Behavior. A paper read to the 1966 American Education Research Association Annual Convention; Chicago, Illinois. February 1966. 6p.

10. Garrison, Harry. A Pilot Study of the Stanford Appraisal Guide of Teacher Competence. Unpublished research report, Seattle, Washington Public Schools: Seattle, Washington; June 1962, p. 1-3.

11. Lorge, I., Fox, D., Davitz, J., and Brenner, M. "A Survey of Studies Contrasting the Quality of Group Performance and Individual Performance," Psychological Bulletin, 55, 1958. p. 337-372.

12. McCall, W. A. Measurement of Teacher Merit. Research report, State Department of Public Instruction, Raleigh, North Carolina, Publ. No. 284, 1952.

13. Medley, Donald M., and Mitzel, Harold E. "Measuring Classroom Behavior by Systematic Observation," Handbook of Research on Teaching edited by N. L. Gage, Chicago, Illinois: Rand McNally & Company, 1965. p. 247-328.

14. Remmers, H. H. "Rating Methods in Research on Teaching," Handbook of Research on Teaching edited by N. L. Gage. Chicago, Illinois: Rand McNally & Company, 1965. p. 329-378.

15. Ryans, D. G. Characteristics of Teachers. Washington, D. C.: American Council on Education, 1960. passim.

16. Wallen, Norman E. and Travers, Robert M. W. "Analysis and Investigation of Teaching Methods," Handbook of Research on Teaching edited by N. L. Gage. Chicago, Illinois: Rand McNally & Company, 1965. p. 448-505.